Statistics Review

Statistics Review

Outline

- Mean
- Median
- Standard Deviations
- Upper & Lower Confidence Limits
- Coefficient of Variance
- Probabilities

Mean

- Defined as simply the average of all the items in a sample.
- 73, 58, 67, 93, 33, 18, and 147.
  - If you added up these values you would get a sum of 489.
  - Divide that sum by 7 to get a mean of 69.9.
- Affected by "outliers", or extreme values.
- Used as basis for determining data variance from the "actual value."

http://www.cmh.edu/stats/definitions/

6 April 2005 NAVOSH PDC

Median

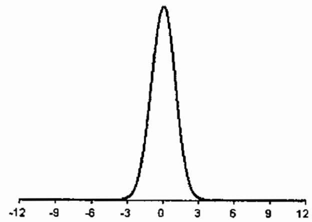
- The value so that roughly half of the data are smaller and roughly half of the data are larger.
- 1.84, 2.96, 3.49, 3.68, 3.72, 3.73, 3.84, 3.84, 4.14, 4.41, 4.80, 4.26, 5.57, and 5.85.
  - For this data there are an even number of observations (**n=14**).
  - Select halfway between the 7th observation (3.84) and the 8th observation (also 3.84).
  - The median is 3.84
- Medians not affected by extreme outliers unlike averages.

http://www.cmh.edu/stats/definitions/

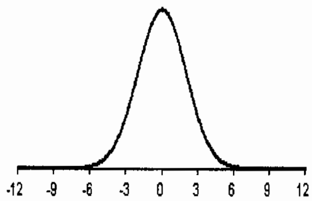
6 April 2005 NAVOSH PDC

Standard Deviation

- Standard deviation is a measure of how spread out your data are.



Standard Deviation of 1



Standard Deviation of 2

http://www.cmh.edu/stats/definitions/stdev.htm

Standard Deviation

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

- n = Number of samples in the data set.
- X_i = Individual Data Measurement
- $\overline{X}$ = Mean of the data set

http://www.cmh.edu/stats/definitions/stdev.htm

Standard Deviation

The 7 values in this data set are 73, 58, 67, 93, 33, 18, and 147.
The mean for this data set is 69.9.

$(73-69.9)^2$	=	$(3.1)^2$	=	9.61
$(58-69.9)^2$	=	$(-11.9)^2$	=	141.61
$(67-69.9)^2$	=	$(-2.9)^2$	=	8.41
$(93-69.9)^2$	=	$(23.1)^2$	=	533.61
$(33-69.9)^2$	=	$(-36.9)^2$	=	1361.61
$(18-69.9)^2$	=	$(-51.9)^2$	=	2693.61
$(147-69.9)^2$	=	$(77.1)^2$	=	5944.41

The sum of these squared deviations is 10,692.87

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

Divide by 6 to get 1782.15

Take the square root of this value to get the standard deviation, **42.2**.

http://www.cmh.edu/stats/definitions/stdev.htm

Coefficient of Variance

- The degree to which a set of data points varies.
- Different from a Standard Deviation:
  - Measures the "preciseness" between 2 or more sets of data
  - Normally given as a percent
- CV = 100 (SD/$\overline{X}$)

Example:

Noise Meter A measures machine 1, 10 times, with mean of 105 dB and std dev of 20 dB.

CV = 100 (20/105) = 19%

Noise Meter B measures machine 1, 10 times, with mean of 97 dB and std dev of 30 dB.

CV = 100 (30/97) = 30.9%

Here, Noise Meter A is more precise than Noise Meter B.

Upper and Lower Confidence Limits

• The interval estimate gives an indication of how much uncertainty there is in our estimate of the true mean.

• The narrower the interval, the more precise is our estimate.

• OSHA Compliance for full-period, continuous single air sample is based upon a 95% confidence level

<u>1st Step</u>: Divide exposure by Permissible Exposure Limit to determine standardized concentration

$Y = (X/PEL)$ where X is the air sampling limit, PEL is the OSHA Permissible Exposure Limit

<u>2nd Step</u>: Calculate the UCL and LCL

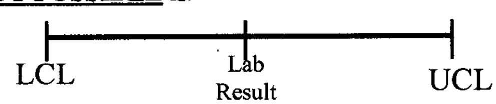
UCL = Y + Sampling & Analytical Error (SAE)

LCL = Y - SAE

Upper and Lower Confidence Limits

• SAEs: Represent a variation due to sampling and errors in analysis
 - • Combined effects in sampling pump, the analysis in the lab, and the pump flow.
 - • These values are normally given to you (per OSHA reference).

• UCL represents the upper most limit of a estimated value
• LCL represents the lowest limit of a estimated value
• OSHA Violation **<u>DEFINITELY</u>** Exists if:
 - • $LCL \geq 1$
• OSHA Violation **<u>MAY EXIST</u>** if:
 - • $LCL \leq 1$ and $UCL > 1$
• OSHA Violation **<u>NOT POSSIBLE</u>** if:
 - • $UCL \leq 1$



Upper and Lower Confidence Limits

An OSHA officer takes a full-period, continuous single sample of a substance, whose PEL is 130 parts per million. The lab analysis records a value of 134 ppm, but according to OSHA references, the lab/sampling error is 0.1. Is there an overexposure at this workplace? What are the 95% confidence limits?

1st Step: Divide exposure by Permissible Exposure Limit to determine standardized concentration

$$Y = (X/PEL) = (134\ ppm/130\ ppm) = 1.03$$

Here, 1.03 > 1.00 so therefore, there is a possible overexposure

2nd Step: Calculate the UCL and LCL

$UCL = Y + SAE = 1.03 + 0.1 = 1.13$

$LCL = Y - SAE = 1.03 - 0.1 = 0.93$

Conclusion: 95% confident that an overexposure MAY EXIST.

6 April 2005 NAVOSH PDC

Upper and Lower Confidence Limits

An OSHA officer takes a full-period, continuous single sample of a substance, whose PEL is 130 parts per million. The lab analysis records a value of 100 ppm, but according to OSHA references, the lab/sampling error is 0.1. Is there an overexposure at this workplace? What are the 95% confidence limits?

1st Step: Divide exposure by Permissible Exposure Limit to determine standardized concentration

$$Y = (X/PEL) = (100\ ppm/130\ ppm) = 0.76$$

2nd Step: Calculate the UCL and LCL

$UCL = Y + SAE = 0.76 + 0.1 = 0.86$

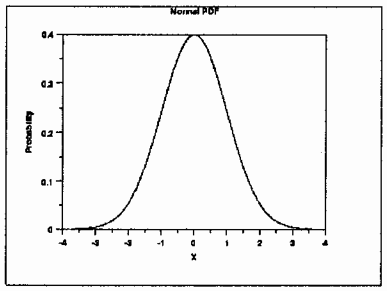
$LCL = Y - SAE = 0.76 - 0.1 = 0.66$

Conclusion: 95% confident that an overexposure DOES NOT Exist.

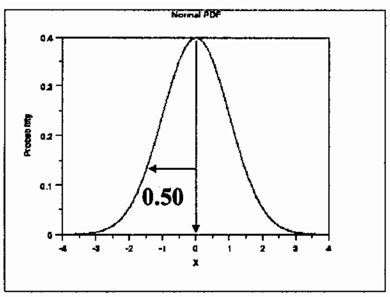
6 April 2005 NAVOSH PDC

Probabilities

- A number expressing the likelihood that a specific event will occur.
- Must use the Z tables in the BCSP Formula Sheet
- Probabilities are mostly based on a normal distribution curve.
  - The Z-values in the Z-table represent an area underneath a part of the curve.



Probabilities

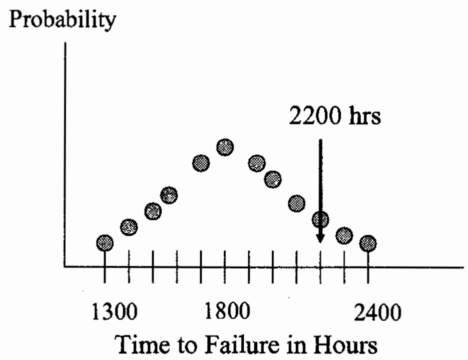


1st: Determine Z:
$Z = (X - \bar{X})/SD$

2nd: Look at the value of Z in the Normal Distribution Table. That value represents the area of the curve from that Z-point to 0.

3rd: The probability is determined by totaling the z-value with the rest of the distribution curve (e.g., 0.50).

Probabilities



A braking system was found to fail as depicted in the graph with a mean of 1800 hours of use. A Standard Deviation of 250 was found based upon the set of data. What is the probability of the component failing at 2200 hours of use?

1st: Determine Z:
Z = (X – X)/SD = (2200-1800)/250
= 1.6

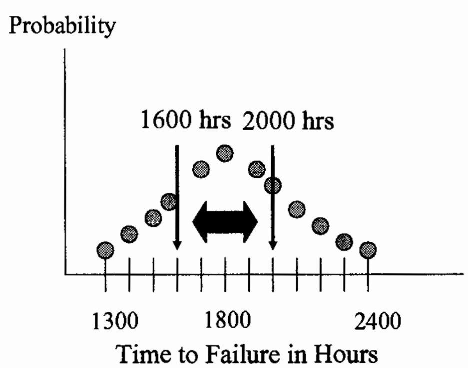
2nd: Look at the value of Z in the Normal Distribution Table.
Z = 1.6 has a curve of 0.4452

3rd: The probability is determined by totaling the z-value with the rest of the distribution curve (e.g., 0.50).

0.50 + 0.4452 = 0.9452 = **94.5% probability of failure at 2200 hours of use**

6 April 2005 NAVOSH PDC

Probabilities



A braking system was found to fail as depicted in the graph with a mean of 1800 hours of use. A Standard Deviation of 250 was found based upon the set of data. What is the probability of the component failing between 1600 hours of use and 2000 hours of use?

1st: Determine Z for each time frame:
Z = (X – X)/SD
= (2000-1800)/250 = 0.8
= (1600-1800)/250 = -0.8

2nd: Look at the value of Z in the Normal Distribution Table.
Z = 0.8 has a curve of 0.2881
Z = -0.8 has a curve of -0.2881

3rd: Here, you are interested in 2 areas: positive side of the peak, & negative side of the peak. Therefore, you ADD the Z-values:

0.2881 + 0.2881 = 0.5762 or 57.6% probability of failure between 1600 & 2000 hrs of use.

6 April 2005 NAVOSH PDC

Poisson Distribution

• Poisson distribution arises when you count a number of events across time or over an area.

• You should think about the Poisson distribution for any situation that involves counting events.

• Conceptually, one determines whether events occurring is a poisson distribution vice a normal standard distribution is if:

1. The **probability of observing a single event** over a small interval is **approximately proportional to the size of that interval.**

2. The probability of **two events** occurring in the same narrow interval is **negligible**.

3. The **probability** of an event within a certain interval **does not change over different intervals.**

Poisson Distribution

$$P = \frac{\lambda^x e^{-\lambda}}{x!}$$

P = Probability
λ = Average of the past occurrences
x = Estimated future number of occurrences

You are a safety supervisor for Factory Fantastic. In Factory Fantastic, you average about 4 accidents during a given month. What is the probability that there will be 6 accidents in the upcoming month?

$$P = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^6 e^{-4}}{4!} = \frac{(409.6)(0.018)}{4 * 3 * 2 * 1}$$

= 0.3072 = 30.7% chance that there will be 6 accidents in the upcoming month.

Poisson Distribution

$$P = \frac{\lambda^x e^{-\lambda}}{x!}$$

P = Probability
λ = Average of the past occurrences
x = Estimated future number of occurrences

In previous year, aircraft major accident rate was 1.53 accidents per 100,000 hours of flying time. As a flight safety officer at a Naval Air Station, were you have 25,000 hours of flying time total per fiscal year, what is the probability that your base will experience exactly 2 aircraft major accidents this year ?

Determine your average: λ
25,000 hours x (1.53 accidents/100,000 hrs) = 0.383 = λ

$$P = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(0.383)^2 e^{-(0.383)}}{2!} = \frac{(0.147)(0.682)}{2 * 1}$$

= 0.050 or 5.0% chance that there will be exactly 2 major aircraft accidents this year.

Poisson Distribution

$$P = \frac{\lambda^x e^{-\lambda}}{x!}$$

P = Probability
λ = Average of the past occurrences
x = Estimated future number of occurrences

On your large Army Post, the base police report that an intersection on 1st and Chestey Puller Drive has an average of 6 accidents per month. Due to budget constraints,however, the base commander will only allow you to take action, if there is the probability that more than 3 accidents will occur in that intersection 50% or greater. What is theprobability that more than 3 accidents will occur this month?

1st: When you see the word "more", you must determine the individual probabilities of an event occurring less than the number of occurrences concerned.

$$P_0 = \frac{6^0 e^{-6}}{0!} = 0.0025$$

$$P_1 = \frac{6^1 e^{-6}}{1!} = 0.015$$

$$P_2 = \frac{6^2 e^{-6}}{2!} = 0.045$$

$$P_3 = \frac{6^3 e^{-6}}{3!} = 0.089$$

$P_{<3}$ = 0.152 or 15.2% probability that 3 or less accidents will occur at this intersection this month

Therefore, the probability that MORE THAN 3 accidents would occur at this intersection this month is
1 - 0.152 = **0.848 or 84.8%** probability that more than 3 accidents will occur this month at this intersection.

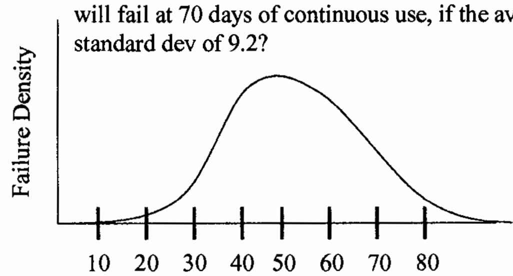
Group Exercise

1. As the Safety Manager, you have 7 generator stations, all of which are identical in terms of their power output and manufacture make/model. Your IH staff measure the following noise levels: 95 dB, 100 dB, 97 dB, 102 dB, 99 dB, 110 dB, 105 dB. What is the standard deviation of the noise generated?

2. Your response staff has two Photoionization Detectors for detecting leaks in and around your storage tank farm. PID A measures 15 different locations in the tank farm and gets a mean of 5.6 ppm with standard deviation of 3.2. PID B measures the same 15 locations as PID A, but gets a mean of 4.6 ppm with standard deviation of 4.2. Which PID has better precision?

Group Exercise

3. You noticed that hydraulic system of your 1K forklifts have failed after constant use. You tracked the failure rates below. What is the probability that you 1K forklift will fail at 70 days of continuous use, if the average failure rate is 48 days, with standard dev of 9.2?



Group Exercise

4. Centers for Disease Control and Prevention reports a incident rate of 3.5 per 100,000 of food poisoning in the US last month due to food eaten/purchased at a food service establishment. In your base, where you have over 20 snack bars/concessionaries leader's clubs serving food, what is the probability that your base of 50,000 will get more than 2 cases of food poisoning this month?

Practical Test
Statistics & Probability Review

5 Questions
10 Minutes to Complete

Battle Drills
Statistics & Probabilities

1. Data was collected of sound level measurements in a indoor firing range: 110 dB, 103 dB, 120 dB, 111 dB, 105 dB, 117 dB. What is the mean of these measurements?

   a. **120 dB**
   b. **111 dB**
   c. **105 dB**
   d. **103 dB**

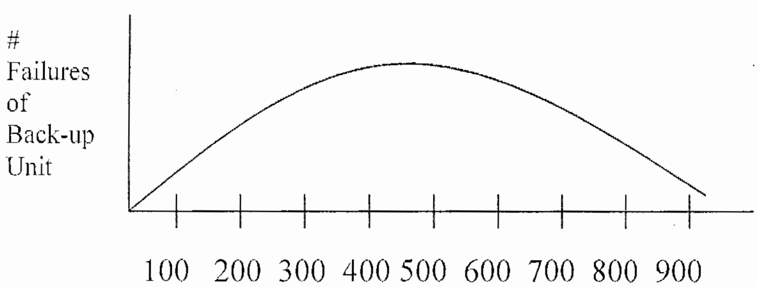
2. Using the data set in problem#1 above, calculate the standard deviation.

   a. 5.46
   b. 6.60
   c. 2.22
   d. 1.34

3. Your safety office was asked to evaluate two light meters. Evaluating different office and warehouse spaces, you listed 2 sets of data with 20 data points. Light Meter A had a mean of 153 foot-candles for all of the spaces measured with standard deviation of 3.43. Light Meter B had a standard deviation of 5.32 with a mean of 176 foot-candles for all of the same spaces. Which light meter would you recommend for your superiors?

   a. **Light Meter A**
   b. **Light Meter B**

4. As a safety officer for Ace Forklifts, Inc, you have been tracking the failure rates of the back-up units for your 1-K forklifts. Based upon the graph below, a mean failure rate at 450 hours of use, and a Standard Deviation of 175, what is the probability that the back-up unit will fail at 700 hours of service?



   a. 75%
   b. 78%
   c. 88%
   d. 92%

5. As the base safety officer, you read an annual service-wide report that the active duty death rate due to private motor vehicle accidents is 5.5 per 100,000. For this year, then, what is the probability that you will have exactly 1 active duty death due to a private motor vehicle accident this year if you have an active duty population of 2500?

a. 12%
b. 23%
c. 7 %
d. 5 %